Web Accessibility Initiative

Web Content Accessibility Guideline – A step towards inclusion

Shilpi Kapoor[#]
[#]*BarrierBreak Technologies*
[1]*shilpi@barrierbreak.com*

ABSTRACT— The World Wide Web can assist people who are marginalised such as Persons with disabilities to live an empowered life. The Web Content Accessibility Guideline (WCAG) is such a step towards inclusion of these people. This paper will discuss the need and the impact of WCAG and Accessible Rich Internet Application (ARIA) in the India Scenario.

KEYWORDS— Web Content Accessibility Guideline, WCAG, Persons with Disabilities, Accessible Rich Internet Application

I. INTRODUCTION

The World Wide Web has helped to break many barriers. It has made people from round the world come together.

Persons with disabilities are one such group that have been largely marginalised in India. The web can give them a new lease on life. The Web Content Accessibility Guideline (WCAG) is a step towards inclusion of these people. This paper will discuss the need and the impact of WCAG and Accessible Rich Internet Application (ARIA) in the India Scenario.

II. WEB CONTENT ACCESSIBILITY GUIDELINE

The Web Content Accessibility Guideline (WCAG) 2.0 has been developed keeping in mind achievability and applicability.

A. Impact of Accessibility

WCAG considers the needs of different types of disabilities, for example visual, hearing, physical, speech, cognitive and neurological, multiple disabilities, and age-related conditions.

As per a study conducted on the Wide Range of Abilities and Its Impact on Computer Technology [1], with regards to difficulties and impairments among Working-Age Adults:

- Approximately one in four (27%) have a visual difficulty or impairment.
- One in four (26%) have a dexterity difficulty or impairment.
- One in five (21%) have a hearing difficulty or impairment.
- Somewhat fewer working-age adults have a cognitive difficulty or impairment (20%)
- Very few (4%) have a speech difficulty or impairment.

In addition, the WCAG also consider the need of the aging population.

Another area that also needs a mention is the overlaps between the Mobile Web Practices and WCAG, also make it important for us to consider the WCAG guidelines and its impact.

B. WCAG Explained

WCAG 2.0 has a total of 12 guidelines that are organized based on the 4 principles, that are:

- Perceivable
- Operable
- understandable
- robust

WCAG 2.0 works on the concept of testable success criteria. These have been divided into three levels, namely:

- Level A
- Level AA
- Level AAA

C. India Scenario

The Guidelines for Indian Government Websites [2] have adopted WCAG 2.0 Level A guidelines. Thereby we are harmonizing our standards with WCAG 2.0.

The National Portal of India (http://india.gov.in and http://bharat.gov.in) as well as the Ministry of Social Justice and Empowerment website has been made accessible keeping in mind Level AA compliance of the WCAG 2.0.

III. ARIA

As the lines between the web and applications are reducing, as content is getting more and more dynamic in nature, we have to consider the needs of persons with disabilities.

It has always been assumed that if the website has to be used by a visually impaired it should be a text only website. This myth is soon breaking.

Today we see the use of technologies such as AJAX, Javascript, etc on the rise. For the same W3C has worked on the Accessible Rich Internet Applications (W3C-ARIA).

As per ARIA, we need to add attributes to identify the role, the current state and the how elements relate to each other. We can now facilitate navigation on a web page by marking regions such banners, menus, primary content or secondary content.

In the Indian scenario, live regions in W3C-ARIA have been implemented in RTI Online (http://rti.india.gov.in)

IV. CONCLUSION

It is upto the developers and the delivery agents to ensure that websites are accessible to persons with disabilities.

Why would we like to increase the digital divide, especially when technology can create an impact in the lives of persons with disabilities and empower them in the day-to-day life.

The future of India needs to be that every citizen of India can access the Indian government as well as private corporation websites. That a citizen can apply for his/her own passport or file his/her income tax return online.

A step towards Inclusion of persons with disabilities in main stream life.

REFERENCES

[1] Forrester Report, *The Market for Accessible Technology—The Wide Range of Abilities and Its Impact on Computer Use*, 2003.

[2] *Guidelines for Indian Government Website*, http://web.guidelines.gov.in

The Challenges in Designing Web Browsers for Visually Impaired People

Ritwika Ghose, Tirthankar Dasgupta, Anupam Basu
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
West Bengal, India
ritwika.ghose@gmail.com, iamtirthankar@gmail.com, anupambas@gmail.com

ABSTRACT— The World Wide Web is an excellent medium through which access to a vast amount of knowledge and information has been possible very easily and practically at no cost. However, the visually handicapped people have difficulty in accessing web contents as most web sites have accessibility barriers, like graphical representations, multimedia, etc. In this paper we present a brief review of two aspects of the problem. First, we present a detailed study on the document structure of different types of HTML pages, their characteristics and the underlying difficulties in accessing contents from those pages. We also look at navigability issues for diverse types of web pages found in the www. Finally, we present a novel architecture of an open source, light weight web browser that seeks to meet the above challenges. The proposed architecture allows a blind person to navigate any web content through simple speech commands and voice feedback to any keyboard operation. The browser will have an integrated text extraction engine that inspects the content of the page to construct a structured representation. This helps in easy and flexible navigation of the page so as to rapidly home into objects of interest. The browser is integrated to an automatic Text-To-Speech and Text-To-Braille transliteration engine that outputs the selected text in the form of speech and/or Braille.

KEYWORDS— Web Browser for blind; types of web pages; content reorganization; text-to-speech; text-to-Braille

I. INTRODUCTION

Over the years, the World Wide Web (WWW) has been an excellent medium for a number of benefits of mankind. It has facilitated communication between people located in any part of the world. It has provided access to unlimited information in various genres. Apart from a number of services, numerous kinds of entertainment factors can also be found on the Web. Though all these facilities have been extremely profitable to mankind, sadly, the visually handicapped people face great difficulties in accessing these benefits. Further, the unstructured nature of different pages along with multi-modal information restricts a blind user to access contents of a web page easily. Thus, in order to provide proper information access and to bridge the communication gap between the visually impaired and the sighted community, the need to build some advance technologically supported systems that will allow a blind person to access web contents easily and efficiently is indispensable.

A number of attempts have been made in the past to overcome this problem, like screen readers [1,2,3,4], speech enabled web browsers [5,6,7,8,9,11], and automatic Braille transliteration systems [2,3,12,13]. However, a number of aspects of web surfing like, navigability issues, presentation of the pages in a structured fashion, and recognition of speech commands have not been addressed so far.

In this paper, we have tried to overcome the above mentioned challenges by proposing a novel architecture of an open source, light weight web browser for the blind people that can be used to access web contents and perform web based tasks more easily and efficiently. To achieve the above mentioned task first, we have presented a detailed analysis on the document structure of different web pages available in the WWW. Our analysis shows that the presentation of each type of page has to be different according to the need. Yet, a general structure is needed to organize the parts of the page and portrayed in a user-friendly manner.

The rest of the paper is organized as follows, in section II we present our detailed survey of the related works in this field and the aspects of web browsing for the blind, which have been attempted previously. A study on the different types of web pages and their structures has been put forward in section III. In section IV, we have proposed a new architecture of a web browser for the visually impaired, which attempts to solve the above problems. Finally, section VI concludes the paper and gives an idea about our future works.

II. RELATED WORKS

In this section we will discuss about some of the previous attempts that have been taken to build web browsers for the visually challenged people.

The basic requirement of web surfing for the visually handicapped people is presenting the text output of the pages into speech. The early works in this field consisted of such screen readers. For example, ASAW from MicroTalk, Screen Reader/2 from IBM [1] were some of the screen readers developed. More recent screen readers consist of HAL screen reader [2] developed by Dolphin and JAWS [3] for Windows from Freedom Scientific. These screen readers also provide Braille output. The disadvantage of simple screen readers is that they just present the web page as it is in the form of audio. It does not provide any kind of interpretation of the text resulting in a monotonous reading of the text. Emacspeak [4] is another speech interface which was designed to provide the corresponding underlying information along with the text.

Screen readers cannot address the problem of easy navigation of the users. It is necessary for a visually handicapped user to listen through the whole text, read out by a screen reader from top to bottom, to get an idea about what is present in the whole page. A sighted user can directly read their area of interest as they get an idea about the page at a glance. Also, some contents like images, graphics are of limited use to a visually handicapped person. Some web browsers have been developed in order to address the above problems. The WebbIE browser [5] extracts the text, uses alt names for images, or removes them, and represents the html page as a simple text, making it easy for a screen reader to read it. Moreover it allows skipping of links and highlights headings for better understandability of the user. [6] describes a browsing solution aims to structure the web pages for ease in browsing. However, due to the complexity and diversity of web pages, all kinds of pages could not be simplified. [7] presents another web browser which generates a tree structure from the HTML documents through analyzing links. Though this attempted to structure all the pages linked together, it did not prove to be efficient for surfing. Moreover, navigability and usability issue of the current page was also not handled. Another browser that was developed to deal with the problems was eGuideDog [8], which had an integrated Text-To-Speech engine. This system applies some advanced text extraction processes to represent the web page in a more user-friendly manner. Another speech output browser, called BrookesTalk [9] aims to address the aspect of efficient understanding of the page by the user. However, both of these still did not meet the standards of commercial use. [10] describes an accessibility Kit for blind people using a new language Blind Markup Language (BML). IBM's Home Page Reader [11] used a numeric keypad for web surfing. Links and normal texts were read in different gender voices so that clear distinction of the two types can be understood. ShrutiDrishti [12], a web browsing system covered some new aspects like conversion of text to Braille. Users can browse the web sites with single key input /minimum key combinations and provides a very user-friendly interface. BrailleSurf [13] also translated the content of the screen to Braille. This browser implemented simplification, i.e. removal of images, etc., rephrasing, i.e. adding labels before links, restructuring the page and use of different reading strategies, like extensive reading, or reading only links.

III. ANALYSIS OF WEB PAGES

The nature of web pages varies widely across the Internet. Apart from content, the structure of pages may differ quite a lot depending on what is to be presented to the user. The size and constituent parts of the page depend on the type of page and its content. For example, some pages may have more graphical images while some may not need them at all.

For a visually handicapped user, it is necessary to present the pages in audio or tactile format. Less important features like images or animations may be skipped. However, just reading the text aloud serially will not be efficient as the user has to listen through the whole session to be acquainted with what the entire page offers.

Also, it is essential to present the different types of pages in different ways, as the flow of information is distinct in each type. Since the number of such types is huge, a generalization of the rules of presentation has to be obtained. In order to achieve that, we now present a survey of some broad categories of web pages.

A. Single Article Page

This kind of a page keeps the focus on a single main theme. An example of this kind of page is Wikipedia[14]. The basic DOM structure of such a page is illustrated in Figure 1.

B. Multiple Article Page

This kind of a page presents articles about more than one topic of almost equal importance in the same page. Only a part of the full article may be presented for each topic.

C. Search Engine

This type of pages takes in an input from the user and returns some result like a number of links to pages related to the input keyword. For example, Google is a search engine. Search option may also appear as a part

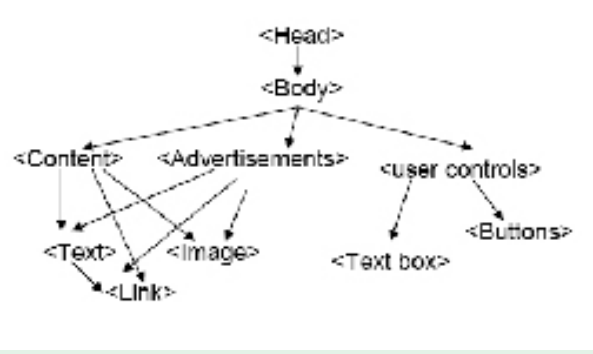

Fig. 1. Illustration of the DOM structure of a single article page.

of a page, and the rest of the page may contain other information. Example of this category is Yahoo.

The basic structure of a search engine, before search is as follows:

- Text area to enter input string
- Search button.
- Category lists. (Optional)
- Link to advanced search options. (Optional)
- Few other non-related links. (Optional)

The basic structure of a search engine after obtaining the result is as follows:

- Links to pages related to the search string
- Small information about where the keyword lies in page corresponding to the links.
- Related search strings. (Optional)
- Links to other non-related pages.
- Advertisements

D. Email

The structure of an email page is very different from the other pages. The whole page consists of different types of links.

E. Brochure Sites

These sites usually describe a business, a person, a company, a place, etc. Its main aim is to provide information about itself to promote awareness. The number of pages for these sites is usually less, thus reducing the number of links.

F. Retail Sites

The retail sites are online shops, from where people can buy any merchandise. This mainly consists of information about the goods that are available, customer help, etc.

G. Social Networking and Community Sites:

These sites are very popular over the Internet. These consist of user's personal information, pictures, places to interact with people and many other options. This kind of a page is slightly different in structure than other types of web sites. It usually consists of a number of links.

H. Forums and Blog

These pages require the user to write something.

I. Online Songs and Videos

These are websites which offer the users to play songs and videos online. These offer user operations like play, pause, next, stop, etc. It usually does not contain much information.

J. Web Based Forms

These pages also come as a part of any web site which requires the user to fill a form. Structure of the page consists of the normal html structure, with text boxes; submit button, radio and check boxes, etc.

With this study on the broad types of pages, we can have general structure of the page. This is shown in figure 2.

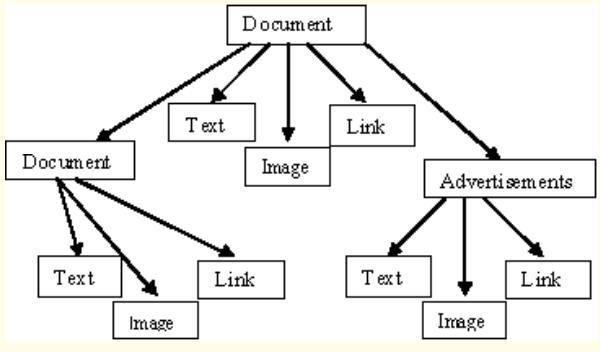

Fig. 2. Generic Structure of a web page

Based on this study of web pages, we have proposed an architecture, which utilizes this generic structure of web pages to re-structure each page. This organized page can then be used to present to the visually handicapped user to achieve better navigability. This architecture is described in the next section.

IV. PROPOSED ARCHITECTURE

We have proposed an architecture of an open source web browser that makes it easy for the visually handicapped

people to browse the web. It contains three modules:

(1) User Input Module,

(2) Text Extraction and re-organization, and

(3) Output Representation.

1) User Input Module:

The user input module of the web browser deals with the different input methods. The system has two types of input.

i) Speech input, passed through an automatic speech recognition engine, and converted to corresponding browser commands.

ii) The other one is keyboard input with an optional voice feedback.

2) Text Extraction and Re-organization Module:

The structure of the web page, that is desired to be presented, is first extracted using an HTML parser. Then the syntactic information is extracted from the tags. Then this information can divide the page into different sections and then presented to the user in a convenient format.

The generic structure of the page thus obtained can then represent the page in a hierarchical fashion. Headings will represent the top levels of the hierarchy. Less important portions like advertisements will be given least priority. This structure is then used to decide the order of presentation of the contents of the page to the user.

Alternate names can be used in place of images and graphical representations. If that is not available, these will be dropped.

Information about links is also presented to the user so that he/she gets a clear picture of the web site and where it can lead to.

3) Output Module:

The output module of the web browser has two key features.

i) Text-to-Speech engine, which converts the modified text of the page into speech output.

ii) Text-to-Braille Transliteration engine, which applies a transliteration rule to the text document and converts it into Braille. This can then be printed using a Braille Embosser and obtain a Braille document.

CONCLUSION

The presentation of a web page ideally varies with the type of its content. Some generic types of web pages have been discussed in this paper. The architecture of a web browser for the blind should be able to capture the type of web page that is being presented and vary its output accordingly.

ACKNOWLEDGEMENT

The authors would like to thank Department of Science and Technology, Govt. of India for financially supporting this project and the members of Communication Empowerment Laboratory for an excellent environment for this work.

REFERENCES

[1] Jim Thatcher, Screen Reader/2 Access to OS/2 and the graphical user interface.

[2] http://www.yourdolphin.com/productdetail.asp?id=5

[3] http://www.freedomscientific.com/products/fs/jaws-product-page.asp

[4] T.V. Raman, Emacspeak —Toward The Speech-enabled Semantic WWW.

[5] King, A., Evans, G., & Blenkhorn, P., WebbIE: a web browser for visually impaired people. In S. Keates, P. J. Clarkson, P. Langdon, & P. Robinson (Eds)., Proceedings of the 2nd Cambridge Workshop on Universal Access and Assistive Technology, Springer-Verlag, London, UK, 2004

[6] Chu, C.: Two Dimension Interactive Voice Browser for the Visually Impaired. In: K. Miesenberger eds. ICCHP 2004, LNCS 3118. Berlin: Springer, (2004), 721–724.

[7] Liang, Z., Song, X., Zhu, Z., Liu, R.: Design and Implementation of a New Browser for Blind Person. Computer Engineering and Applications. (2004), 14, 106-108.

[8] Jing X., GuanNeng H., and Yong T., An Open Source Web Browser for Visually Impaired, LNCS 4681, pp. 90–101, 2007. © Springer-Verlag Berlin Heidelberg 2007

[9] Mary Zajicek, Chris Powel, Chris Reeves, A Web Navigation Tool for the Blind.

[10] Mercedes Macías, Fernando Sánchez, Improving Web Accessibility for Visually Handicapped People Using KAI, Proceedings of the 3rd International Workshop on Web Site Evolution (WSE.01)

[11] Chieko Asakawa, Takashi Itoh, User Interface of a Home Page Reader.

[12] http://medialabasia.in/index.php?option=com_content&task=view&id=100&Itemid=149

[13] www.dinf.ne.jp/doc/english/Us_Eu/conf/csun_99/session0158.html

[14] www.wikipedia.org

Darpan - Simplification and Relevant Usage Models A Case for enriching computer usage.

Sachin Kelkar, Jagadish Babu
Intel technology India Pvt Ltd
Sachin.kelkar@intel.com Jagadish.babu@intel.com

ABSTRACT— India's PC penetration is relatively low compared to other developing countries. India's Total market size of PC connected to Internet is about 9 million devices[A]. At the same time, mobile subscriber base has cross 500 million. Though Personal Computers have become a part of daily life and its services are being increasingly used by consumers across India, its widespread adoption for personal use is yet to happen.

Use of Personal computers is limited to technology savvy English literate users. The research done by Intel team showed that there is a need to simplify the PC user experience and this will directly result in more usage of PC and Internet. Use of local language for content Internet content consumption and availability of relevant usage models will increase the Internet and PC penetration in the country.

To validate these assumptions, Intel team worked for 8 months to create a simplified UI interface and create an ecosystem of software and content partners. The fundamental design principles of this simplified UI initiative (Project Darpan) were defined as

- **Simplified UI to take away perceived complexity in Computing**
- **Create relevant usage models for Indian consumers for PC + Internet**
- **Use Local language to provide access to India relevant content**
- **Create a framework for monetizing for content providers**

The paper will discuss the assumptions and data pertaining to this research, field activities to validate these assumptions, Challenges faced while developing the Simplified UI framework and opportunities for the ecosystem in following this approach. The paper will also share examples of how this project is being used as a proof of concept in some of the large PC penetration campaigns across the country.

KEYWORDS— Personal Computers, Simplified, User Interface, Local Language

I. INTRODUCTION

Darpan, initiative was formed to explore ways to improve usage of computing devices among masses and strategize methods to remove barriers that prevent usage. In this paper, we intend the briefly introduce the problem statement of usage and briefly present the current strategies to stimulate the usage for PC and Internet from an Indian Consumers.

The base research for this study came from various internal market research done by Intel Corporation and particularly research done by Internet and Mobile Association of India (IAMAI).

It is very articulated fact that the number of mobile subscribers in the country has crossed the 500 Million mark. However, in the same market the number of Personal Computers owned by consumers are about 9 million. However while the owners of Personal computers may be 9 Million, there are over 71 Million people who use internet and computers in some form, supposedly using shared computers. It was interesting as to why only 9 million users thought of investing to buy a PC while the the rest chose to use shared resources. Interesting dependency on language literacy emerge which hings on English awareness of the user population.

II. DEMOGRAPHY OF INDIA

This population is spread across different socio economic classes and speaks and read different languages. Does their non familiarity with English blocks them from even using technology tools such as PC and internet.

India is a country of many languages. It could be safely inferred that, existing computers and computer interfaces clearly require Knowledge of English.

Figure 1: Shows for the purposes of this study, figurative analysis of Indian population bifurcating based on literacy, and English language awareness, further divided based on urban and rural distribution.

Research indicates that 63% of urban population and 83% of rural population are not familiar with English. Could the lack of awareness of English a major reason for less usage overall?

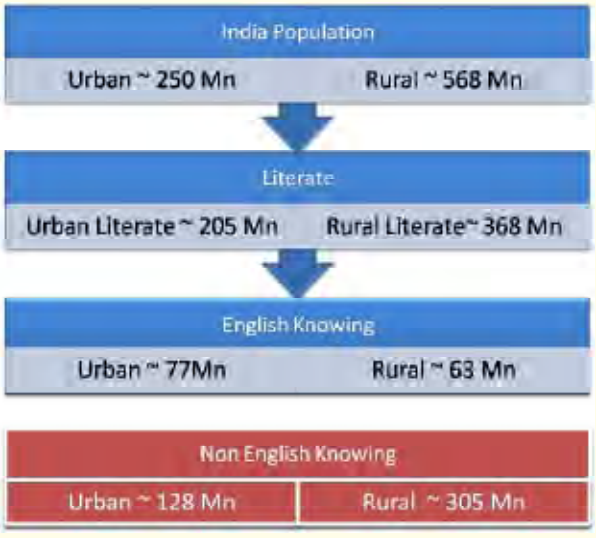

Fig. 1. A population distribution model for India in Urban/ Rural and English literate Vs Non literate

Out of ~ 77 Mn English knowing urban population , the PC literate pollution is anticipated to be ~ 62 Mn. This establishes clear correlation between adoption of PC and comfort level of English. Apart from access and affordability , another deterrent for the adoption of computers is the absence of local language computing and content. To spread the PC awareness and consumption, breaking the language barrier is very essential.

In order to further refine the targeted audience, dividing the user base along usage proficiency/Usage of computers was found useful. Figure 3 depicts the division along existing usage. PC Literacy term is coined to indicate understanding of basic working nature of PC and having basic usages understood. Attributes were developed to describe the users. These attributes in parts or sum total depicts the user categories. Further it should be added that these attributes may not be comprehensive.

PC Literacy	Category	Attributes
Literate	Power User	Owns a PC, Uses it very often. English educated. Savy Users.
Recent Literate	Explorer	Recent Owner, or User but not owner.Could be fluent in English/Vernacular language, May have been introduced to computers at work / home / Vocational courses.
Illiterate	Potential near Future User-1	Literate. May know English. Or may know only local languages not English. May never been introduced to PC at schools/work. May have fear of technology. May not have seen value to his lifestyle.
Illiterate	Potential Future Users-2	Potential far out user. Illiterate. May be already a cell phone owner. May not simply be able to use PC in its existing forms. Needs Innovations to address needs as well as methods of access.

Darpan Initiative focuses on Enhancing usage could be potentially focussed on Explorer and Near Future categories .

A further breakdown of factors affecting the usage disparity, study focussed on enhancing further usage of existing users and understanding the needs of non users.

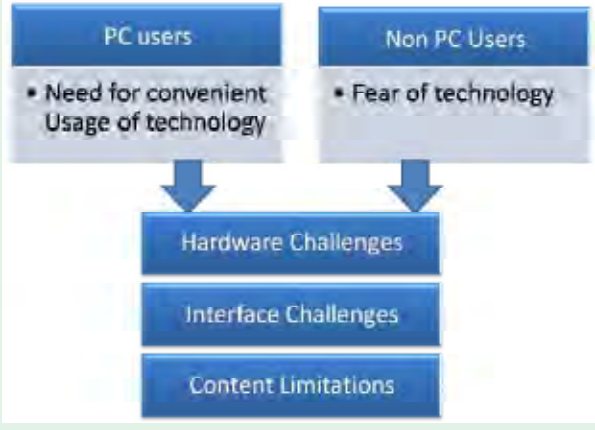

Fig. 2 Barriers for Personal Computers Usage

Figure 2. Breaks the problem among three large categories. Fear of technology, lack of familiarity with user interface and finally lack of relevant usage models for persistent usage.

III. DARPAN THOUGHT PROCESS AND MARKET POTENTIAL

Darpan focussed on understanding these challenges and develop a novel and innovative approach which will make Personal computing devices a integral part of an Indian house hold. Darpan focuses on developing interventions to ease barriers experienced at each stage and for various categories of users.

Traditional personal computers, characterised by extensive use of Keyboards and mouse, Ionized and menu driven user interfaces that need users to develop good on screen navigational skills using sensitive mouse and finally keyboards which have non sequential letter

arrangements, may appear all too familiar for the proficient users. Through formal/informal education, existing users have understood the finer details of using personal computers.

In order to understand the challenges posed by Interface, Darpan hypothesized that Simpler interface, characterized by large ICONS depicting usages (not applications) ,could potential develop larger usage.

Further on, Internet as understood by most of proficient users needed good understanding of what internet is, and how one would use it. One would need to understand the terminology of internet such as a address of websites and meaning of search, hyperlinks so on so forth. It is amazing to understand the speed with which these nuances are understood and assimilated by newer categories of users.

Darpan further hypothesized that the user interface

- that never requires user to type address in the address bar,
- provides for magazine like experience with most of the commonly used categories of articles/web content,
- accessible using clicks, six button interfaces

Will significantly boost simpler usage of Internet.

In the IT sectors, Indian local language usage has been significantly enhanced over years. With development of Font layouts and innovative translation/transliteration techniques, Users are able to interact / develop content in local languages. For obvious reasons, users tend to like to interact and consume content in their local languages. One of the research articles suggest that over 70% of internet users prefer to consume local language content. [1]

Darpan team first explored internet for local language content. Through transliteration technology, many current users were exchanging / developing well developed content on Internet. However, it was also found that use of native keyboard layouts are very complex to learn and needs lot of training before they can be used proficiently.

Darpan further hypothesized that local language content consumption is far higher priority for the user than letting users develop their own content. As an example, reading a Hindi web page may be far higher on their list of usage as compared to writing email in Hindi. Accordingly, the simple user interface in local language was hypothesized to enhance usage across vernacular sections of our populations. Technology evolution may result in efficient translation technologies, simpler inputting techniques can create significant content generation.

Usage of any device is dependent on types of applications and services available on the device. In order to create further sticky usage models, darpan team proposed that a new ecosystem of application/ content developers be created who will develop content according to darpan's other principles especially around simplified local language interfaces for use with various computing devices from Personal computers, handphones and smart TVs.

It was hypothesized that without these sticky usage models, users may not find internet interesting enough and may be deterred by complexity of using the interent.

In order to validate these hypothesis, intel teams conducted extensive market research including end user testing/pilot activities/

While it is obvious to see that persistent usage may lead to eventual ownership. A possible sequence of interventions leading to usage and eventual ownership is depicted here in the 3.

The size of the opportunity was derived at by the basic research and the estimated potential was arrived at by estimating the number of house holds which can potentially buy a PC. It was estimated that with this approach, there is a great opportunity for PC market to grow in India.

IV. PROOF OF CONCEPT – DESIGN AND VALIDATION

We used the Intel Software Partner Ecosystem to identify possible partners to engage on this initiative. We zeroed in on M/s GoDB Tech Pvt Ltd to create an early Proof of Concept solution to validate the assumptions made earlier . The early POC was available in October 2009 and was tested across existing PC channel partner communities and through focused group discussion with consumers across India. The field research validated the assumptions of Ease of Use , Relevant usage models and Local language and provided valuable insights in the consumer's wish list for content and usage behaviour.

Some of the feedback during the research showed that the categories of the purposes preferred by the consumer are as followed :

V. PRODUCT DESIGN AND ROADMAP

The Proof Of Concept was then evolved into a software & Content framework design which allows easy

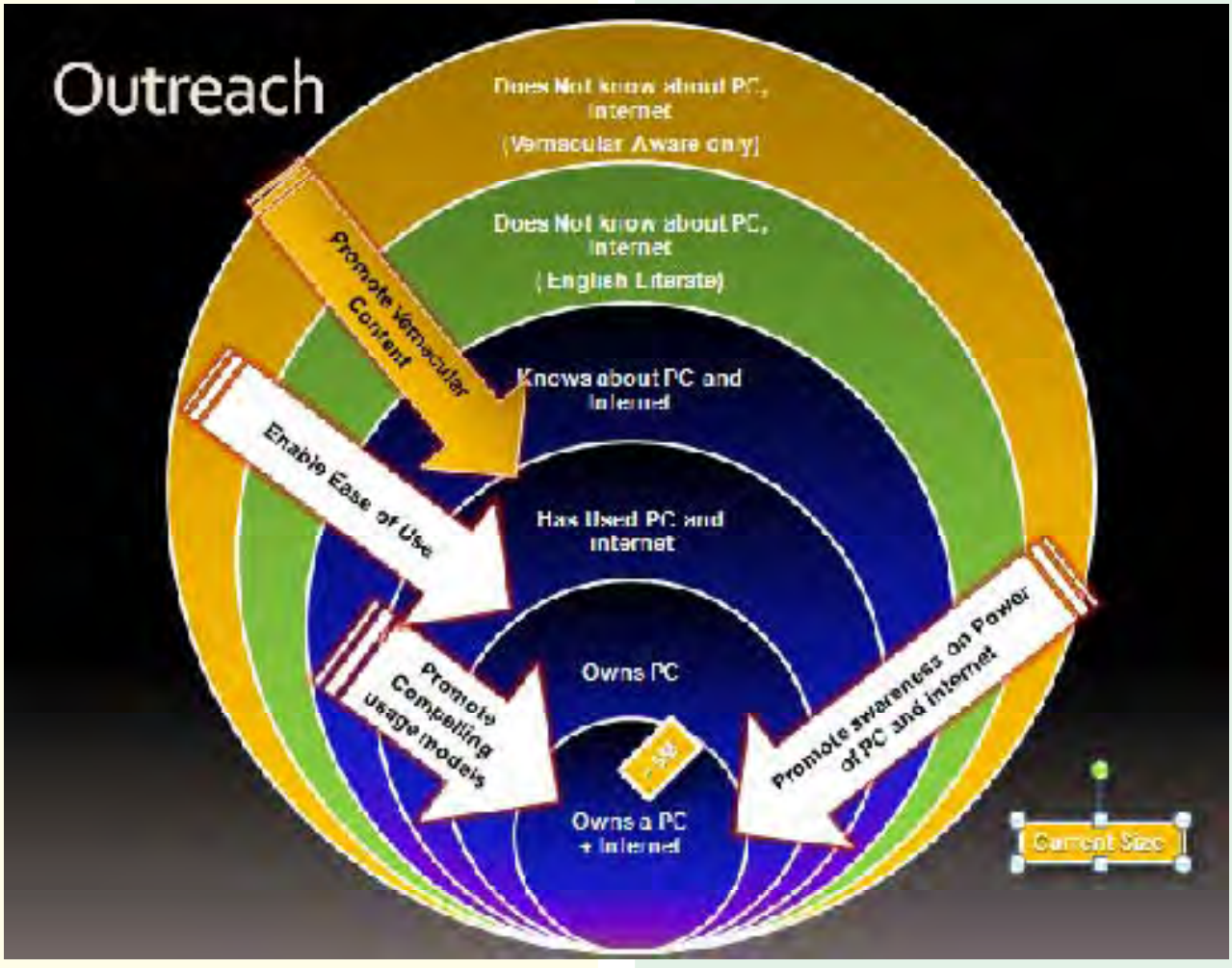

Fig. 3 Market Opportunity with *Darpan initiative*

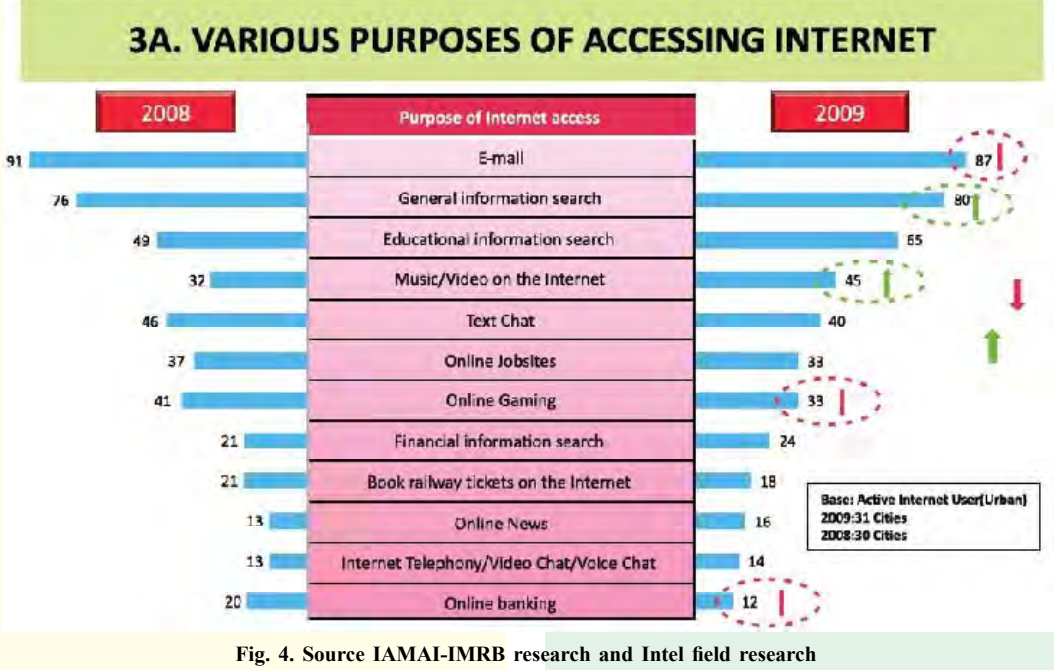

Fig. 4. Source IAMAI-IMRB research and Intel field research

integration of content feeds from various sources. A detailed product roadmap was conceived to address the 4 vectors of ***Darpan*** initiative .The focus of the Content Framework and UI framework was restricted to content consumption in the first phase and challenges of content creation to be tackled in the subsequent phases.

The roadmap also defined the local language integration plan and the likely content partnerships. The framework allows integration of existing internet content using a set of rules and using industry standards protocols such as rss feeds .

The framework will allow multiple content providers to be part of this initiative and provide their content in a simplified format to the consumers. The roadmap also defines a plan to monetize the premium content and provide revenue sharing opportunities with the content and software partner ecosystem.

We are also exploring evolution of multiple UI frameworks which will eventually address specific usage models such as Government Services , Education and Farmers PC etc.

Other area we are exploring pertains to Hardware design simplification. However, since it is in its early stage, most of the details pertaining to bag of ideas and its validation is in its infancy. We would look forward to discuss those in details in coming months.

VI. GO TO MARKET STRATEGIES

It was decided that the User Interface under the ***Darpan*** Initiative to be made available free of cost to all the potential users. The distribution mechanism is being worked out and will be available through a variety of sources in near future. Currently the interface can be downloaded from www.darpan.me .

The Darpan initiative will create an ecosystem of content providers and innovative software companies which will impact the consumption of internet content in a big way , thus by fuelling a local language content ecosystem spiral.

This initiative led by Intel will create a robust business models for all the ecosystem players and will benefit the end consumer immensely.

Fig. 5. Interface Menu screen shots

Fig. 6. Interface Menu screen shots

The early adoption of Darpan initiative has been done in large programs like " MAZA PC " in Maharashtra and we expect this to be used in many large initiatives in the future.

ACKNOWLEDGMENT

We wish to acknowledge valuable contribution from M/s GoDB Tech Pvt Ltd and other Software and content partners who made this initiative possible in a quick time. We also wish to acknowledge the Sales and Marketing team at Intel India who had contributed immensely in the validation and the Go to Market planning of this initiative.

REFERENCES

[1] Vernacular Content Market in India – A Study by IAMAI- IMRB

[2] I-Cube 2009 , 2010 , A study by IAMAI- IMRB

[3] Bhashaindia.com by Microsoft India

[4] Intel CMR Research on Consumer Market

[5] State of local language computing – Various publications from http://tdil.mit.gov.in/

[6] "We are like that only - understanding the logic of consumer India" by Rama Bijapurkar

Accessibility of all Information to all people

Veera Raghavendra and Kishore Prahallad
International Institute of Information Technology, Hyderabad.
raghavendra@iiit.ac.in kishore@iiit.ac.in

ABSTRACT— Our paper focus on issues involved in accessibility of information of types: speech, text, to all people which include literate, illiterate, visually challenged and differently abled persons. We discuss the issues involved, and provide suggestions or recommendations of required standards to be developed for various language technology components involved in human-machine interaction over web. We believe many such standards are not available as on date, and development of such standards aid in building better interfaces both on desktops as well as on mobile to provide accessibility of all information to all people.

KEYWORDS— Speech-to-Speech, ASR, TTS, Machine Translation, Information retrieval.

I. INTRODUCTION

Now-a-days people are highly depending web for their day-to-day life. The users of the web are literate, illiterate and visually challenged people. Literate users can manage the web without any issues. But, illiterate and visually challenged users require a literate help. The goal of the web is to facilitate communication between people who speak different languages and due to the increasingly globalizing world economy, humanitarian services and national security, there is an ever increasing demand for speech-to-speech translation. While substantial progress has been made over the past decades in each of the related areas of Automatic Speech Recognition (ASR), Machine Translation (MT), Information Retrieval (IR), Summarization, Natural Language Processing (NLP), Text-to-Speech (TTS) system. W3C is so far created standards in the areas of ASR and TTS. More standards are required for utilizing IR, MT and summarization and more over global standards are required to utilize them altogether.

The rest of the paper is organized as follows. Section 2 describes the each component elaborately. In Section 3, available W3C standards for various components are discussed. Section 4 discusses what standards are required for each component.

II. COMPONENTS OF SPEECH-TO-SPEECH

Automatic Speech Recognition

Speech recognition is the process of converting a speech signal, captured by a microphone or a telephone, to a set of words. Speech recognition system can be characterized by many parameters. An isolated-word speech recognition system requires that the speaker pause briefly between words, whereas a continuous speech recognition system does not. Spontaneous, or extemporaneously generated, speech contains disfluencies, and is much more difficult to recognize than speech read from script. The conventional statistical framework employed to accomplish the speech recognition comprises three major components – acoustic models, language model, and the pronunciation dictionary. Speech recognition applications include voice dialling, call routing, content-based spoken audio search, simple data entry, and medical transcription.

Machine Translation

Machine translation is a automated process of converting the text from one language to another. For example to Telugu text to Hindi text or English text to Hindi text. To process any translation, human or automated, the meaning of a text in the original language must be fully restored in the target language. Translation is not a mere word-to-word substitution. A translator must interpret and analyse all of the elements in the text and know how each word may influence another. This requires extensive expertise in grammar, syntax (sentence structure), semantics (meanings), etc., in the source and target languages, as well as familiarity with each local region. Current machine translation software often allows for customization by domain or profession (such as weather reports) — improving output by limiting the scope of allowable substitutions. This technique is particularly effective in domains where formal or formulaic language is used. Human and machine translation each have their share of challenges. For example, no two individual translators can produce identical translations of the same text in the same language pair, and it may take several rounds of revisions to meet customer satisfaction. But the greater challenge lies in how machine translation can produce publishable quality translations.

Information Retrieval

The techniques of storing and recovering and often disseminating recorded data especially through the use of a computerized system. Automated information retrieval systems are used to reduce what has been called "information overload". Many universities and public libraries use IR systems to provide access to books, journals and other documents. Web search engines are the most visible IR applications. Two main approaches are matching words in the query against the database index (keyword searching) and traversing the database using hypertext or hypermedia links. Keyword searching has been the dominant approach to text retrieval since the early 1960s; hypertext has so far been confined largely to personal or corporate information-retrieval applications.

Summarization

Summarization is the restating of the main ideas of the text in as few words as possible or in a new, yet efficient, manner. There are different types of summaries depending what the summarization program focuses on to make the summary of the text, for example generic summaries or query relevant summaries (sometimes called query-biased summaries). Summarization systems are able to create both query relevant text summaries and generic machine-generated summaries depending on what the user needs. Summarization of multimedia documents, e.g. speech, pictures or movies are also possible. Some systems will generate a summary based on a single source document, while others can use multiple source documents (for example, a cluster of news stories on the same topic). These systems are known as multi-document summarization systems.

Text-to-Speech System

A Text-to-speech system deals with conversion of text into spoken form. Now-a-days, TTS systems are used in many applications such as car navigation systems, information retrieval over telephone, voice mail, language education, screen readers, speech-to-speech translation systems and so on. The goal of a TTS system is to synthesize speech with natural human voice characteristics and, furthermore, with various speaker specific individualities and emotions. TTS system comprise of mainly two components; text analysis and waveform generation. Text analysis includes dividing the text into sentences and words, assigning syntactic categories to words, grouping the words within a sentence into phrases, identifying and expanding abbreviations, recognizing and analysing expressions such as dates, fractions, money, and grapheme-to-phone conversion. The second component is generally referred to as a synthesizer which generates the speech waveform for the given sequence of phones.

III. W3C STANDARDS FOR VARIOUS COMPONENTS

Speech Synthesis Mark-up Language (SSML)

SSML [1] is responsible for rendering a document as spoken output and for using the information contained in the mark-up to render the document as needed by author. The following are the six major processing steps undertaken by a synthesis processor to convert marked-up text input into automatically generated voice output.

1. XML parse: An XML parser is used to extract the document tree and content from the incoming text document. The structure, tags and attributes obtained in this step influence each of the following steps. A simple English example is "cup<break/>board"; the synthesis processor will treat this as the two words "cup" and "board" rather than as one word with a pause in the middle.

2. Structure analysis: The structure of a document influences the way in which a document should be read. For example, there are common speaking patterns *associated with paragraphs and sentences.*

Mark-up support: The <p> and <s> elements defined in SSML explicitly indicate document structures that affect the speech output. A <p> element represents a paragraph. An <s> element represents a sentence.

```
<p>
   <s>This is the first sentence of the
paragraph.</s>
   <s>Here's another sentence.</s>
 </p>
```

3. Text normalization: The written text may contain non-standard words. Such as numbers, dates, telephone numbers, etc., Text normalization is an automated process of the synthesis processor that performs this conversion. For example, for English, when "$200" appears in a document it may be spoken as "two hundred dollars".

Mark-up support: The **say-as** element can be used in the input document to explicitly indicate the presence and type of these constructs and to resolve ambiguities. many acronyms and abbreviations can be handled by the author via direct text replacement or by use of the **sub** element, e.g.W3C can be written as World Wide Web Consortium.

```
<sub alias="World Wide Web
Consortium">W3C</sub>
```

2. Text-to-phoneme conversion: Word pronunciations may be conveniently described as sequences of phonemes, which are units of sound in a language that serve to distinguish one word from another.

Mark-up support: The **phoneme** element allows a phonemic sequence to be provided for any word or word sequence.

```
<phoneme alphabet="ipa" ph="t&#x259;mei&#
x325;&#x27E;ou&#x325;"> tomato </phoneme>
```

3. Prosody analysis: Prosody is the set of features of speech output that includes the pitch (also called intonation or melody), the timing (or rhythm), the pausing, the speaking rate, the emphasis on words and many other features.

Mark-up support: The **emphasis** element, **break** element and **prosody** element may all be used by document creators to guide the synthesis processor in generating appropriate prosodic features in the speech output.

```
That is a <emphasis level="strong">
huge </emphasis>
```

4. Waveform generation: There are many approaches to this processing step so there may be considerable processor-specific variation.

Markup support: The **voice** element allows the document creator to request a particular voice or specific voice qualities (e.g. a young male voice). The audio element allows for insertion of recorded audio data into the output stream.

```
</voice>
   <voice name="Mike">I want to be like
Mike.</voice>
```

VoiceXML: VoiceXML [2] is an XML language for writing Web pages you interact with by listening to spoken prompts and jingles, and control by means of spoken input. VoiceXML brings the Web to telephones. VoiceXML has been carefully designed to give authors full control over the *spoken dialog* between the user and the application. The application and user take it in turns to speak: the application prompts the user, and the user in turn responds.

VoiceXML documents describe:

- spoken prompts (synthetic speech)
- output of audio files and streams
- recognition of spoken words and phrases
- recognition of touch tone (DTMF) key presses
- recording of spoken input
- control of dialog flow
- telephony control

The following example offers a menu of three choices: sports, weather or news.

```
<?xml version="1.0"?>
<vxml version="2.0">
<menu>
  <prompt>
  Say one of: <enumerate/>
  </prompt>
  <choice next="http://www.sports.
example/start.vxml">Sports </choice>
  <choice next="http://www.weather.
example/intro.vxml">Weather</choice>
  <choice next="http://www.news.
example/news.vxml">News</choice>
  <noinput>Please say one of
<enumerate/></noinput>
</menu>
</vxml>
```

This dialog might proceed as follows:

Computer:	Say one of: Sports; Weather; News.
Human:	Astrology
Computer:	I did not understand what you said.(a platform-specific default message.)
Computer:	Say one of: Sports; Weather; News.
Human:	Sports
Computer:	(proceeds to http://www.sports.example/start.vxml)

VoiceXML integrated speech synthesis, speech recognition and telephone interface together. Similarly, all other components of the speech-to-speech system need to be established and combined together.

IV. REQUIRED STANDARDS FOR EACH COMPONENT

So far W3C created standards for speech synthesis

and speech recognition. To achieve speech-to-speech communication over web, W3C have to come up with some standards for other components. Following are the few required standards for each component.

Text-to-Speech System

This component plays major role in providing accessibility to visually challenged people. Now-a-days many news papers are putting their content over web. But, these vendors use their own fonts and follow their own standards for creating web page. Instead W3C should create standards to these news websites. Few standards are as follows:

1. Font type Vs Unicode: Unfortunately, there exists several popular websites which provide news in their local fonts instead of in Unicode.
2. Structure: Separate tags for headings, short stories, full stories, headlines. Such structure would help in accessing the data for the purpose of TTS, MT and summarization.

Machine Translation

A new and great initiation required for creating machine translation standards similar to SSML or VoiceXML. The standards should address the MT aspects such as.

1. What is the source language?
2. What is the target language?
3. POS information of source language text.
4. Possible word-to-word mappings (In case of foreign words).

Another issue is the integration of MT with TTS and ASR. So far, the integration or APIs to integrate MT, TTS and ASR is specific to a particular implementation party.

Information Retrieval and Summarization

Information retrieval and summarization are useful during speech-to-speech communication. If the user is looking for cross-information retrieval from the web, the system has to convert the text into target language and corresponding information has to be retrieved. To achieve this W3C has to provide some standards. Following are the few standards.

1. Source language key words need to be searched.
2. What kind of documents need to be searched.

DISCUSSION AND CONCLUSION

An important aspect to be considered in further defining the standards is the user type. So far, the standards seem to be mostly applicable to common persons. However, with the role of ICT beings enhancing, differently abled persons, especially visually challenged are accessing the information over WWW. There are a few commercial and academically developed screen reading software which provide accessibility to the information. However, there seems to no standards either in their interfaces or in the way the information is available on the WWW for these screen reading software. For example, a user after getting acquaintance with a particular screen reading software may find difficulty in switching to other screen reading software as the interfaces (keys performing specific function) might differ, Also from the developers perspective, the future of information access lies in multilingual speech-to-speech mode, where the role of MT, IR, summarization play a roles apart from ASR and TTS engines. Sufficient and efficient standards need to be evolved in order to integrate these various components for providing access of information to all people. Yet another direction which is not discussed fully in this paper is multi-modal aspect of multi-lingual speech-to-speech access of information. It would not be difficult to image to use both audio and video in the near future to access information. Hence standards may to have generic to incorporate video mode of information access too.

REFERENCES

[1] Speech Synthesis Mark-up Language, http://www.w3.org/TR/speech-synthesis/.

[2] VoiceXML, http://www.w3.org/Voice/.